

Digitising of Map Archives in the National Land Survey of Sweden

Stefan Gustafsson
National Land Survey of Sweden
S-801 82 GÄVLE
E-mail stefan.gustafsson@lm.se
Fax: +4626 65 40 95

Summary

Digitising of older maps and documents from the archives of the National Land Survey started in the early 1990-ties. Because of an archives strategy founded in 1998 these activities by now have resulted in a decision to go on in a large-scale realisation of digital archives in the organisation. After the digitisation, the original material is to be delivered to the authority responsible for governmental archives, the National Archives of Sweden. The number of deeds or separate maps will exceed 2.5 million. In total, more than 50 million maps and text-pages will be digitised. The digital version will be accessible on Internet for internal and external users.

By end of 2001, already 20% or 10 million items will be digitised, and the present plans are to complete the digitisation by 2005.

The Archives of the National Land Survey of Sweden

History, Contents, Localisation

The archives of the National Land Survey take their origin from the beginning of the organised surveying activities in Sweden. In 1628, the mathematician Andreas Bureus was charged the task to make up maps over a large part of the villages of Sweden. The focus was on land owned by peasants, liable to taxation.

At the time, the king, Gustaf II Adolf, was engaged in the Thirty Years' War, and this, together with his ambition to establish firm Government authorities stimulated the foundation of the first surveying authority. The maps, called geometric land books (geometriskas jordeböcker) as they were collected in bindings, were to supplement the earlier written land books.

From the beginning, the material was archived in the royal castle in Stockholm, but was moved to a new building in the beginning of the 1690-ties. This saved the material from being destroyed in the great fire in 1697 when the entire castle was burnt down.

From the middle of the eighteenth century, the central archives were provided with material from the redistribution activities aiming to modernise the agriculture of Sweden. The redistribution era continued more than one hundred years with several different recurrent activities. From that time, also regional archives were set up, one in each county, serving the regional survey organisation. At a redistribution activity, three acts containing maps and descriptions were set up. The original, the decision deed, was archived in the regional archive. Two copies were made from that; one of them was delivered to the parties concerned and the other, called "public renovation" to the central archives. This material is also edited so it is easier to take in. The documents sometimes also have an aesthetic more attractive design. All information in the original deeds is however not copied to the central version. Since the 1940-ties, no more information has been brought to this archive from the counties.

In the central archives also printed maps from mid 1800-ies and later are stored.

Volumes

The regional archives contain about 2.4 million acts. Every year the archives are conveyed with an additional 10-15 000 acts. The central archives contain more than 250 000 acts or individual maps. Preliminary calculations estimate the total amount of image files to more than 50 million separate image files, if all material is to be digitised. In the circumstances mentioned it naturally is crucial to

make the right decision when choosing method for data capture, file format for storing, storage media and technique for presentation.

Usage

The originals have been used as the basis for documentation of property division and property formation activities throughout the years, and in reality, still are. One has to consider that the archives contain the complete documentation on the division in real properties in Sweden, and thus is crucial for activities related to this, for example planning, ownership, financing, statistics and taxation. Formally, the Swedish property register is only a comprehensive compilation of the archives' contents. To determine the status of the property division the archived information must be retrieved. Some of the originals are because of frequent use worn and might be difficult to read. The regional archives still are provided with the deeds from present property formation activities and other documents concerning the use of land, as building plans, etc.

The central material has not been used in the same extent. It has also been located in suitable premises from the beginning. The quality is therefore in most cases much better.

In total, about 200-250 000 acts are retrieved from the archives every year. The main use of the archives today comes from property formation activities in the National Land Survey. Other significant user groups are, in order of size, privateers, estate agents and different governmental authorities.

The interest from privateers stands probably for the most evident recent increase. The material is, when known and possible to access in an easy way, very informative for genealogists and local historians.

Digitisation of the Archives of National Land Survey

Earlier digitising development and activities

In the early 1990-ties experiments with digitising of historical maps started in the National Land Survey. Attempts were made with different file formats and compression methods. From those activities also knowledge was gained, especially concerning specifications on image quality, suitable file formats, production conditions, etc.

A couple of production lines, mostly dealing with the printed maps from the middle of the nineteenth century until 1950, were set up and a rather large number of maps from different series were digitised, based on these experiences. These activities finished in 1997. The same year a new digitising activity started in co-operation with the National Heritage Board of Sweden. This activity focused on the material from the redistribution activities in the rural areas in the eighteenth and nineteenth centuries. This material has gained an increasing usage as basis for preservation of the culture heritage and as a source for archaeological investigations.

Parallel to this, a first database in MS Access was developed covering the entire content of the archives.

The results from the digitising work, the image files, are collected on different storage media, tapes of several variants and CD-ROM. Naming of files have differed throughout the years, which have made it rather difficult to retrieve a single file. Put together, all this has made the retrieval of digitised material labour-intensive and expensive.

Archives Strategy

In 1998, archives' strategy was decided in the National Land Survey. The strategy infers:

- A creation of digital versions of the paper based original material in a five years perspective.
- Dissemination of the archived material through Internet based interfaces, and in connection with that, creation of a well adapted organisation to this purpose
- Structure, arrange and register all archives, analogue as well as digital
- Depot storage of the archives and continuously ongoing deliverance to the governmental archives authority or a depot arranged by the National Land Survey

Recent activities

The Digital Historical Maps project

The Digital Historical Maps project, DHM, started 1999 as co-operation between the National Land Survey, the National Heritage Board, the National Survey and Cadastre of Denmark, the University of Greifswald in Germany and ESRI Sweden within the EC programme INFO2000. The consortium's overriding objective was to enable professional planners, the multimedia industry, and the public to access and use the historical maps. The aim was also to lay a foundation for the design of databases of archived material as well as specifications and standards for digitalisation, geocoding, and distribution by the Internet.

Within the project user interfaces and cataloguing systems have been developed and tailored to user needs. Moreover, procedures for charging and billing, as well as contracts and procedures relating to copyright issues have been developed.

The DHM project was finished in March 2001 and the results from the project have in great extent influenced the coming solutions for dissemination of historical maps in the National Land Survey. The results from the project can be seen on the project's homepage, www.dhm.lm.se. On the homepage a small selection of maps from the archives of the participating organisations are presented. The homepage will be accessible at least during 2001 and 2002. Later the information probably will be transferred to environments that are more definite on the homepages of participating organisations.

The Digital Archives project

The Digital Archives Project started in 1999 to implement the archives' strategy. The project is divided in three phases. The first phase during 1999 and early 2000, including a preliminary study, ended up in a decision to start the second phase. This second phase included amongst others development of techniques for production and storage of digitised material, methods for archiving of databases, plans for digitising of the analogue archives and archiving direct from digital case management systems. Much focus was given a large scale pilot project for digitising of the entire archives from the regional archives in Stockholm county. As a part of the second phase, an automated archive function is developed that makes it unnecessary to print out copies from the property formation management system.

The third phase of the Digital Archives project will start during 2001 and focus on the digitising of all the archives of NLS. A special unit for preservation work combined with digitising facilities and probably a storing capacity is going to be established in Ånge community. Depending on financial resources, the production time will be from three to six years for the entire archives of NLS.

By now, the production within the project is continuing. By end of 2001, approximately 20% of the regional material will be digitised. In February 2001, it was also decided plan for a big scale accomplishment of the digitisation. Co-operation with the National Archives of Sweden has resulted in a proposal for a jointly run production project, partly to be financed within the EC Structure Funds, partly by the NLS, the National Archives and the county council of Västernorrland. After the digitisation, it is discussed to deliver the originals to the National Archives.

Services on the homepage of National Land Survey

The archives of the National Land Survey will in the future be presented in a dedicated application on our homepage, www.lantmateriet.se. Besides possibility for the public to search and view parts of the archived material it will also be possible to order printouts and image files on line. In a special licensed service the entire digital contents, not only from the archives mentioned above, but also from several other archives will be available, as flight indices for aerial photography, land and geographic databases, geodetic data, etc. The first implementation of this functionality is planned to late 2001.

Legal Aspects

Copyright issues

The Bern Convention and the National Copyright Laws

The national copyright law in Sweden is based on the Bern Convention. The Convention can be described as a model law for copyright. The countries, which are members of the Bern Union, have an obligation to create a national copyright law within the frames laid down in the Convention. The Convention contains mechanisms and rules which the respective countries are bound to follow. Therefore, the national copyright laws within the Bern Union have a common structure and terminology.

In the Bern Convention, it is stipulated that the law of the country where a question appears (The Principle of National Treatment) shall govern matters concerning copyright. If for example a work made in Denmark is abused in Sweden, the Swedish Copyright Act shall govern the process against the abuser. The Convention gives a minimum level of protection to every right holder in the Bern Union.

Work in the public sector

The central 'object' for the copyright protection is 'the work'. The national legislator have, in line with the Bern Convention stated that maps shall be regarded as 'works'.

Time restricted protection

The copyright protection is limited in time. The main rule in the EC-union says that the work shall be protected for seventy years.

This rule must be conferred to the constitutional framework, in the first place the Freedom of the Press Act, which regulate what is 'a public document'. Every map in an archive of a Swedish authority is a 'public document'. One document, which has been put in such an archive, has also, in the sense of the Swedish Copyright Act, been published or made public. If the period for protection has ended, the work is, in this respect, in the public domain.

Databases protection

The Directive on the legal protection of databases (Directive 96/9EC Directive of the European Parliament and of the Council of 11 March 1996) lay the basis for the legal aspects of the database protection. The Directive is implemented in Swedish law.

The database protection consists of two parts. One is based on traditional copyright. If data in a database is structured in such a way that the traditional requirements in the copyright law are fulfilled, the database shall have the same protection as other works. The database is a 'work'. The other part is the 'sui generis protection'. This protection is not within the copyright system but is protection on its own. Technically the mechanisms are implemented in the national copyright law (§49).

The sui generis protection is connected with the economic efforts made by the producer. The investments must be substantial. The producer has, with some exemptions, the exclusive right to exploit - 'extraction' and 're-utilisation' of the whole or a substantial part of - the database under a period of fifteen years from the year following the date of completion. The sui generis protection is only valid within the EC. The protection has therefore limited importance in activities on the Internet.

Conclusion

The conclusion is that there is no copyright on maps older than 70 years as such. With reference to the acts of creations involved in the process, the databases are protected by the copyright law. The databases are also a result of substantial investments and are therefore protected by the sui generis protection in the national legislation. It is a legal question whether a single file is protected through the database protection law. Thus, the conditions are to be ruled by agreements based on the national legislation and case law.

Restrictions due to Privacy Protection.

Due to legislation concerning privacy protection, it is necessary to avoid dissemination of personal data on the Internet without restrictions. As the recent acts in the archives mainly consist of deeds from property formation activities, information about privateers is to be found in abundance. This information must be protected, and therefore the information can only be made accessible in a restricted solution.

Financing model and policy for pricing

Financing of production, maintenance and access to information on landscape and property in Sweden is based on a combination of grants and fares. The model is an expression of a governmental commitment as a guarantor for the systems, at the same time as the users, by demands and financing, get influence on the development of the systems. The model implies:

1. Build-up of systems and databases is financed by grants
2. Grants and fares in combination finance management. Management activities include updating, development and adaptation of information and systems to meet the demands from users and connected systems in the world around.
3. Extraction and distribution of information, for example access to information, is financed by fares.

For the historical material this implies:

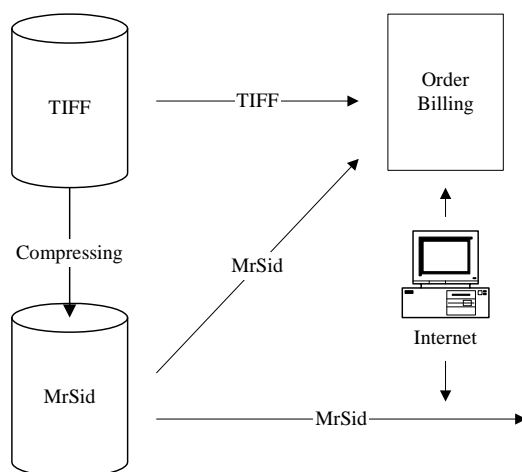
The build-up of databases, including development of the first production system, extraction from archives, indexing, data-capture, occurring techniques for compression and conversion and storage, is to be financed by grants and if possible contribution from other users.

1. Grants and fares in combination will finance management, including the further development of the production system, administration and running of systems. The share of grants is decided in dialogue with the government.
2. Extraction and distribution of information from the databases are financed by fares. For example, if the information is distributed by the Internet, the entire web-solution (web server, the adaptation of information, etc.) is to be financed by fares.

Dissemination Policy

Considering the copyright rules and the financial model a dissemination policy is established in the National Land Survey implying two storage levels and two access levels as follows:

Two Storage Levels



Two Access Levels

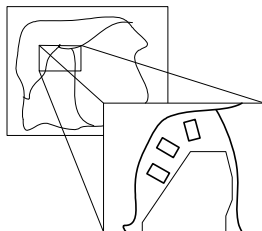
Charged

Entire file (delivered on CD)
-TIFF
-Mr Sid

Free

Browser

Overview map



Detail

The free browser enables

- Overview of entire converted file

- Zoom to detail in limited area with possibility to evaluate quality and make interpretations of the material

- Order of TIFF or converted files

Technical solutions

Image specification

When stating the demands of quality, it is necessary to balance the quality possible to achieve with the financial resources available. Naturally, the result of the digitisation must fulfil the basic needs of reading and understanding the information from the originals. The information must also be able to use for other purposes than presentation. For example, it is necessary to save the images in a format that can be used in other systems for further treatment.

In the National Land Survey, earlier large parts of the archives have been microfilmed in b/w for security reasons. Experiences made from digitising of the microfilm and practical use of the digital version proves that it is possible to create a complete and useful digital version direct from the microfilm. The quality is naturally not on the level possible to achieve by scanning the originals, but with knowledge of the material; it is possible to manage more than 95% of the information research from the digital version from the microfilms. Hereby the production costs will be considerable lower and at the same time, the digitisation will not risk to damage the originals. Not all originals, however, are suited for this method. For example coloured maps and maps with small measurement numbers must be digitised directly from the originals. If the method used in the first production will show not satisfying, the future depots are to be equipped with resources for scanning and printing on demand and experienced personnel for service when needed

It is probably impossible to achieve a true copy of the originals; the digitising process always adds faults into the copy. The ambition is to get a reproduction as close to the original as possible. Although we are trying to set up parameters and requirements that are measurable, objective and can assume a high quality reproduction, where will always be demand for a visual inspection of the results from the production.

The specification for image production used is developed within the National Land Survey of Sweden and based on experiences of digital raster data since 1992.

As the goal for the digitisation is to search and view the image-files in real-time on the Internet, the files also must be very hard compressed and/or presented with a technique that radically reduces the information transfer over the net.

File Formats for Storage

For coloured maps TIFF (Tag Image File Format, ISO 12639) is used. TIFF is the most commonly used format for the raster data interchange and for loss less acquisition. The TIFF format is highly flexible and platform-independent and is supported by numerous image-processing applications.

For text documents the TIFF Group IV is used. The TIFF file format with Group IV compression is used because the act texts are captured mainly from Black & White microfilm, produced earlier for security reasons.

Image Quality

Scanning of coloured maps

The colour system used is the CIE-Lab, a commonly used standard for colour measurement and control. The system can be used for communications and corrections of lightness, saturation and hue. Colour maps are scanned with 24 bits colour depth, RGB Colour Space.

Scanning of Text Documents

Text-originals (Black and white) and B/W microfilms are scanned with one bit colour depth.

Resolution

The resolution of the digital images is set to at least 254 dpi. One pixel is 1/10 mm (50 cm = 5000 pixels) corresponding to the original's size. From experiences in the National Land Survey the 254 dpi resolution gives an acceptable quality and image file size for reproduction of the digital file to a hardcopy in the same size as the original. At the same time, the demands for storage resources are minimised.

Geometric Accuracy

To be able to receive the best geometric accuracy in the final product all steps in the production line have to be set up with demands and requirements of high geometric accuracy. The requirements are set to +/- 0,1%. The tolerance for the accuracy of a map sized 50 cm will therefore be +/-0,5 mm. If the digital data is plotted to a hardcopy, the equipment and paper used naturally also will affect on the result.

Converting tools

There are several tools for compressing TIFF-files for presentation on the Internet. In this area also a continuously development is ongoing. One therefore can expect changes to even better performance in the user interface than today. Having the TIFF files stored as a basis makes it rather easy to change converting tool when needed. Until now, two different tools for converting the TIFF originals have been used, the MrSid and the DjVu software, both distributed by Lizardtech Inc. Both meet our requirements for presentation of the information, but generally, the DjVu format is more effective in compressing. On the other hand, the MrSid format is possible to use in common GIS applications. There are also some differences in the results of the converting. The b/w text documents seem to be better with DjVu and the coloured maps better with MrSid. This is especially the case when the originals contain many colour tones. The examples below are compressed files from scanned original maps. The quality of the images are affected when showed in this document:

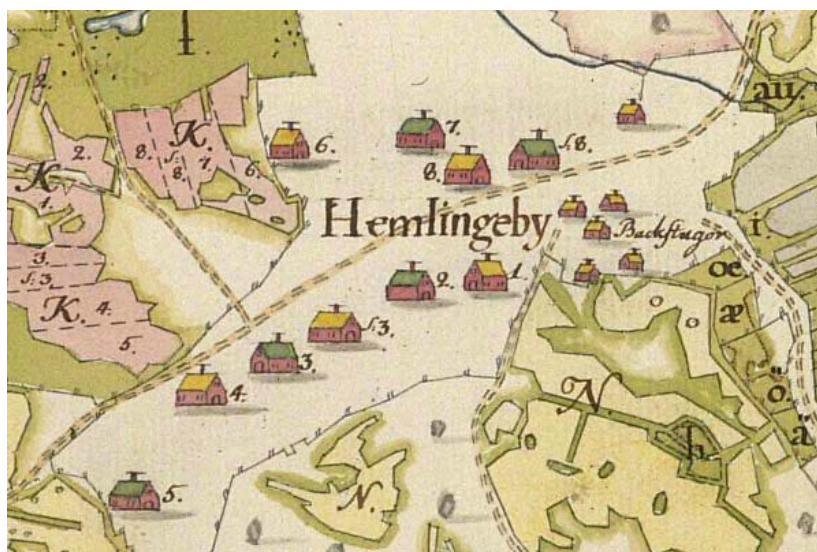


Figure 1. The central archives of the national Land Survey Section contains thousands of beautifully elaborated maps. Map from the redistribution of the village Hemlingby in 1766, Gävleborg county. Mr Sid file sized 8 525 KB, original TIFF file 170 MB



Figure 2. Part of a DjVu-compressed TIFF-image of 424 KB. The original image was 160 MB. The use of this high compression rate makes it still possible to interpret the main part of the information, especially as for the legal contents.



Figure 3. The same image as in figure 7, now compressed with MrSID to 8276 KB. Using this format and corresponding lower rate of compression ensures an almost "true" copy when printed in the original map size.

Examples on text files scanned from B/W microfilm:

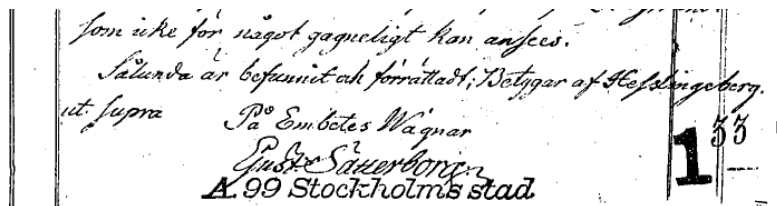


Figure 4. Part of a DjVu-compressed TIFF-image of 178 KkB. The original image was 38,3 MB

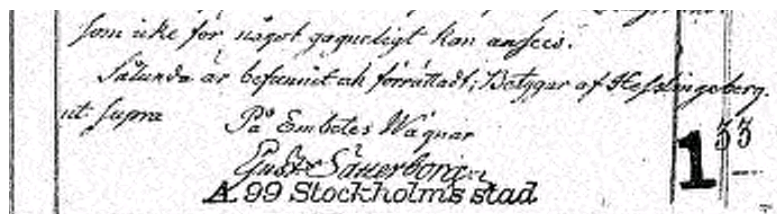


Figure 5. The same image as in figure 1, now compressed with MrSID to the size of 220 KB.

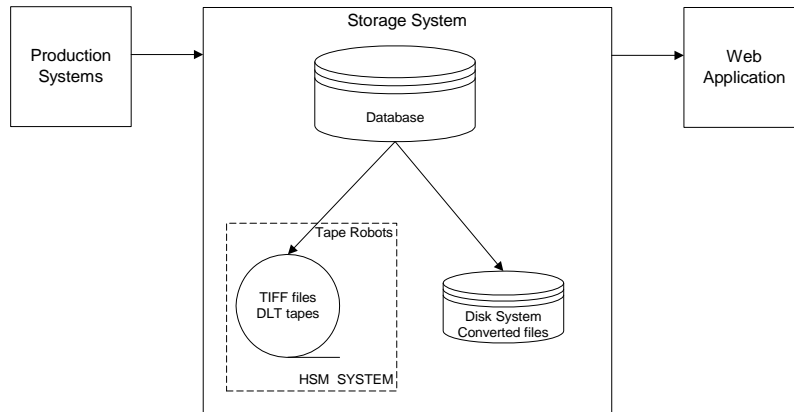
Database solution

In the National Land Survey two different database systems are used for archives management, in

the beginning Oracle, the standard DBMS in the NLS, and lately introduced BRS-Search, especially adapted for large archives systems.

Storage system

The storage used is the standard system for storage of raster data within the NLS:



The storage will hold about 50 million TIFF files when the digitisation is completed. The demand on storage space is estimated to about 120 TB. The TIFF files are kept on DLT-tapes in a hierarchic storage management system, also planned to hold other raster data of the NLS. From the storage, it will be possible to deliver single files on different types of media, as CD-ROM or printouts. From the TIFF storage, it is also possible to convert the files to other formats, when wished.

The storage of converted files will hold about one TB of DjVu files. These are kept on hard disk systems for direct access from the Internet. Parts of the information might be hold as MrSid files to make them directly adapted for different types of standard applications, i.e. PhotoShop, ArcView and MapInfo.

Web Application

The web application will be divided in two functional parts. The first will make not restricted information accessible to the public without charge and make it possible for viewing. Functions for ordering or downloading of TIFF or converted files on payment will be introduced. This part will probably be completed with information on the material, and guide the user how to understand the material. The second part will be a licensed and charged service providing access to the entire material, mainly for professional users.